3 этапа эволюции сетевых файловых СХД

Обзор архитектурных и функциональных особенностей современных NAS-систем компании HP в контексте основных бизнес-трендов на рынке.

Введение

Сетевым файловым системам хранения (или Network Attached Storage – NAS) на базе открытых сетей уже более 10 лет.

Исторически NAS-решения "выросли" из файловых серверов, которые строились на базе стандартных, или серверов общего назначения. Предпосылкой создания NAS-решений была необходимость расширения характеристик обычных файловых серверов (ФС) в корпоративных средах, которые, как правило, поддерживали только один тип протокола доступа и были ориентированы на очень ограниченное число пользователей. В качестве основных требований к NAS-решениям в сравнении с ФС были:

- *гораздо большие масштабируемость, производительность и надежность;*
- *возможность одновременного разделяемого доступа по различным протоколам: CIFS (Common Internet File Service) – Windows-платформы, NFS (Network File System) – UNIX-платформы, а также по протоколам FTP и WebNFS.*

Архитектурно корпоративные NAS-решения строились на основе кластера узлов, представляющих собой независимые файл-серверы (общее число обычно не превышало 20) и использующиеся для параллельного доступа к данным. В качестве непосредственно файлового хранилища в составе решения часто применялись дорогостоящие FC системы хранения. И хотя поставленные перед NAS-решениями задачи выполнялись, ценой расширения их функциональности перед ФС были значительно более высокие как первоначальная стоимость и стоимость администрирования, так и стоимость за единицу информации.

Всю историю "борьбы" за показатели производительности на различных бэнчмарках, протоколах, транспортах, начиная с 1997 г., можно проследить на http://www.spec.org/osg/sfs97.

Разделение рынка в основном на два класса сетевых файловых решений – файловые серверы и корпоративные NAS-решения – первые дешевые, но только для небольшого числа пользователей, вторые дорогие, но с расширенной функциональностью для корпоративных применений – фактически сохранялось до 2006 г. Однако после этого момента, а в отдельных отраслях бизнеса и несколькими годами ранее, стало крепнуть понимание, что справиться с возрастающим объемом файлового хранения, оставаясь в рамках архитектурных решений конца 90-х, становится все сложнее, и требуются уже несопоставимые затраты с имеющимся IT-бюджетом многих компаний. Причиной этому являются глобальные тенденции на рынке.

Основные тренды на рынке сетевых файловых хранилищ

Что говорят аналитики о тенденциях рынка сетевых файловых хранилищ?

- Уже к 2010 г. прогнозируется шестикратное увеличение мировой информации до 988 exabytes в год.
- Вся мировая цифровая информация удваивается каждые 18 месяцев.

Исследования, проведенные ESG, показали, что общий объем постоянных данных к 2010 г. составит более 26 000 Пбайт (рис. 1) при соотношении постоянных данных к "динамическим" как 10:1. Однако уже в ближайшей перспективе это соотношение может возрасти до 100:1.

Тенденции рынка последних лет свидетельствуют о резком увеличении доли неструктурированных данных в общем объеме данных, к которым можно отнести: цифровую графику/видео/аудио, компьютерные модели, результаты компьютерного моделирования, справочную информацию и др. Исследование IDC показывают (рис. 2), что на мировом рынке происходит смещение объемов хранения в сторону файлового. И уже в 2008 г. объем дискового пространства, проданного для файлового хранения превысит объем дискового пространства для блочного хранения.

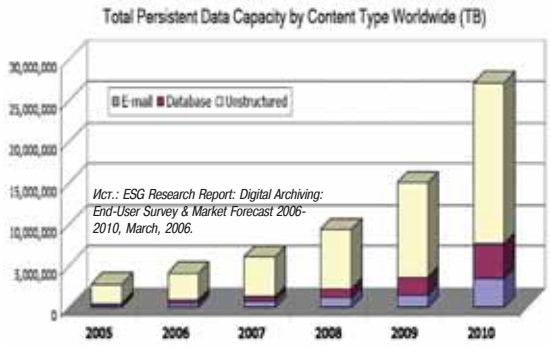

Рис. 1. По исследованиям ESG, общий объем постоянных данных к 2010 г. может вырасти до 26 000 Пбайт.

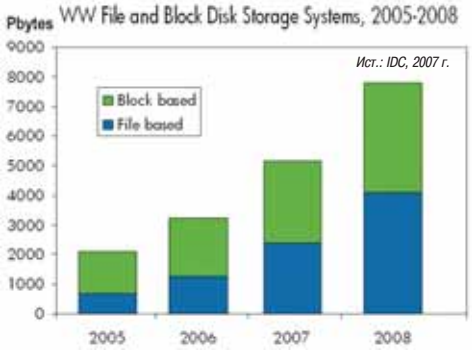

Рис. 2. По исследованиям IDC на мировом рынке происходит смещение объемов хранения в сторону файлового. И уже в 2008 г. объем дискового пространства, проданного для файлового хранения превысит объем дискового пространства для блочного хранения.

Приведем несколько примеров.

Компания Snapfish – крупный провайдер по предоставлению коммунального доступа к частным/общим цифровым изображениям/фотографиям – имея сегодня 6 Пбайт хранимых данных, прогнозирует их рост до 20 Пбайт уже к 2010 г.

Частная компания Aerodata International Surveys (основана в 1992 г. со штаб-квартирой в Бельгии) специализируется на аэрофотосъемке земной поверхности с последующей обработкой снимков и представлением их в виде геоинформационных продуктов. Aerodata использует цифровые камеры с чрезвычайно с высоким разрешением – до 196 мегапикселей для фотографирования земной поверхности. Каждая фотография может занимать до 1,5 Гбайт. За один вылет оцифровывается поверхности примерно на 1 Тбайт данных. Имея 4 самолета, Aerodata может совершать до 6 вылетов в день, что соответствует до 6 Тбайт ежедневно. В 2007 г. Aerodata зафиксировала около 200 тыс. изображений.

При проведении в 2004 г. (г. Афины, Греция) летних Олимпийских игр ежедневно генерировалось более чем 250 000 цифровых изображений (со средним размером 18–24 Мбайт) в течение 17 дней.

Общей особенностью этих примеров является то, что, помимо необходимости хранения больших генерируемых объемов данных, требуются еще и высокопроизводительные каналы доступа (в том числе и коллективного пользования на основе Web 2.0) по записи/чтению к этим данным.

В связи с этими тенденциями бизнес стал выдвигать к файловым хранилищам ряд расширенных требований:

- снижение стоимости на единицу хранения как первоначальной, так и эксплуатационной;
- обеспечение простой многократной масштабируемости системы как по емкости, так и по производительности по требованию в онлайновом режиме;
- существенное повышение управляемости системы с точки зрения объемов хранения на одного администратора, увеличивая ее с единиц терабайт до петабайта и многих петабайт уже в ближайшей перспективе;
- снижение энергопотребления на единицу хранения и на единицу площади, повышая также и конструктивную плотность упаковки системы.

Реакцией HP на требования рынка стало анонсирование в июне с.г. о своей новой системе хранения – StorageWorks 9100 Extreme Data Storage System (ExDS), а также объявление в сентябре с.г. нового интегрированного решения для хранения данных – HP 4400 Scalable NAS File Services, которые сегодня уже доступны на рынке и способны хранить до одного петабайта данных.

Эти 2 решения архитектурно схожи и реализованы на технологии HP Scalable NAS, которая, в свою очередь, строятся на базе ПО HP PolyServe Clustering File System. Первое решение в сравнении со вторым имеет меньшие стоимость в расчете на единицу хранения и функциональность, связанную с файловыми сервисами. Второе – повышенную целостность данных и функциональность, но и более высокую цену на единицу хранения, которая, тем не менее, остается до 50% ниже традиционных корпоративных NAS-решений на рынке.

Необходимо заметить, что новые NAS-решения, упомянутые выше, не следует путать с решениями для архивирования неструктурированного контента (примером котором может служить HP RISS), основными функциями которых являются длительное хранение/архивирование на основе правил информации и быстрый ее поиск в целях удовлетворения регламентирующих требований/законодательных актов. Задачи снижения стоимости хранения, производительности важны, но не приоритетны.

HP StorageWorks 9100 Extreme Data Storage System

ExDS в настоящее время это флагманский продукт семейства NAS-решений HP. Данное решение предлагается в рамках широкой программы развития распределенных вычислений, или “cloud computing”[1)], или т.н. горизонтально масштабируемых сред, которые, помимо этого, активно позиционируются для решений, связанных с услугами предоставления потокового видео, постобработки (post-production) кино-, видеоматериалов, создания цифровых архивов на основе аналоговых носителей, поддержки web 2.0 файловых приложений и др. Для подобных применений могут создаваться центры обработки данных с сотнями и даже тысячами серверов с необходимостью хранения петабайтов информации и удовлетворения самых высоких требований к производительности и энергоэффективности. Для развития этого направления в компании HP было организовано специальное подразделение – “масштабируемые вычисления и инфраструктуры (HP Scalable Computing & Infrastructure)”.

Система ExDS9100 по сути представляет собой горизонтально масштабируемый файловый сервер с оригинальной архитектурой. Если мы имеем дело с обычным файловым сервером, то его можно разделить на 2 части: управляющую (ЦП, ОП, ОС с файловой системой, адаптеры с интерфейсами подключения хостов/внутренних подсистем) и подсистему хранения (RAID-контроллер, жесткие диски). В ExDS9100 эти две части разделены и представляются двумя компонентами (рис. 3), которых может быть ограниченное множество. Первая реализована блэйд-серверами – ExDS Server Blades (на базе шасси HP c-Class BladeSystem серии c7000) – с возможностью масштабирования – от 4 до 16 серверов, вторая – оригинальным блоком хранения – ExDS Storage Unit. Состав базовой и максимальной конфигураций представлен в табл. 1.

Табл. 1. Базовая и максимальная конфигурация ExDS9100

Компонента	базовая	макс.
42U Rack Cabinet	**1**	**2**
c7000 Blade Enclosure	**1**	**1**
Processing Blocks	**4**	**16**
BL460c Blades		
8GB or 16GB RAM (opt)		
Super Saber Mezzanine		
Storage Blocks – 246/820 drives	**3**	**10**
Spitfire – 12 drives		
Freighter – 70 drives		
SAS Switch	**2**	**2**

Базовая конфигурация системы (рис. 3) состоит из четырех блэйд-серверов с многоядерными процессорами и 246 Тбайт (3 блока хранения) для онлайнового хранения. Каждый блэйд-сервер обеспечивает потоковую производительность в 200 Мбайт/с. Как блэйд-серверы, так и емкость хранения могут масштабироваться независимо (как увеличение емкости, так и добавление новых блэйдов выполняется вручную). Блэйд-серверы могут масштабироваться от 4 до 16, емкость хранения – до 820 Тбайт (10 блоков хранения), обеспечивая плотность 12 Тбайт на “U”. При этом общая потоковая пропускная способность системы будет составлять 3,2 Гбайт/с.

[1)] “Cloud computing” представляет собой подход, когда IT-инфраструктура (вычислительные ресурсы и ресурсы хранения данных) доступны пользователю через web-интерфейс в виде сервиса.

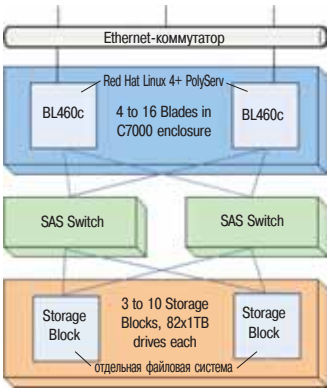

Рис. 3. Структурная схема системы ExDS9100.

Каждый storage-блок включает 82 SAS-диска в 7U рэке: восемь 8+2 RAID6 групп (~78% полезной емкости, защита от отказов двух HDD) с дисками по 1 Тбайт и сдвоенные SAS-контроллеры. Между собой блэйд-серверы и блоки хранения связываются через SAS-свичи, которые конструктивно установлены в шасси серии c7000 (одновременно использующее для установки и блэйд-серверов), что не требует никаких дополнительных коммутаторов.

На каждом блоке хранения – одна или несколько файловых систем. Между собой блоки хранения и их файловые системы не связаны. Каждый блэйд-сервер соединен с каждым блоком хранения и имеет доступ к любой из файловых систем. Запрос по записи/чтению, приходящий на ExDS9100, может быть обработан любым блэйд-сервером, что реализуется специальным алгоритмом выравнивания запросов по блэйд-серверам. Это дает возможность максимально снизить риски появления узких мест на уровне блэйд-серверов, и при равномерной интенсивности распределения запросов по блокам хранения приблизиться к максимальной потоковой пропускной способности.

На каждом блэйд-сервере устанавливается ОС Red Hat Linux 4 и файловая система PolyServe. Если требуются какие-либо файловые сервисы по поддержанию доступности данных (бэкапирование, репликация, моментальные снимки и др.), антивирусное ПО и др., то они приобретаются отдельно в составе соответствующего ПО и устанавливаются на блэйд-сервер.

Система ExDS9100 относится к NAS-рынку, который в настоящее время, как уже упоминалось, в основном разбит на 2, далеко стоящие друга от друга, сегмента: дешевые, низкопроизводительные, немасштабируемые устройства, ориентированные на домашние офисы и намного более дорогие устройства enterprise класса, которые предлагают высокую производительность и масштабируемость, но которые также обычно требуют специальных навыков (сертификации) управления, и гораздо более дорогие. ExDS9100 призвана занять промежуточное положение между ними, т.е. предложить более дешевые NAS-хранилища корпоративного класса.

ExDS9100 позиционируется как система, дающая возможность предоставления многопетабайтной емкости при многомерной масштабируемости "по запросу" и максимальной управляемости ресурсов хранения, позволяя значительно снизить стоимость и упростить управление большим объемом данных.

В результате оригинальных конструктива и архитектуры удалось достичь следующих показателей (*по состоянию на конец ноября 2008 г., прим. ред.*):

- стоимости за 1 Гбайт используемой емкости в составе RAID 6 – менее $2 с учетом всех инфраструктурных компонент, включая блэйд-серверы;
- плотности упаковки в блоке хранения – 12 Тбайт/U;
- плотности упаковки серверов – до 12,8 ядер/U;
- потребляемой мощности – 200 Вт на блэйд-сервер, 497 Вт на шасси c7000, 2070 Вт на блок хранения;
- уровня управляемости объемом хранения в расчете на одного администратора – 1 Пбайт.

Помимо отмеченных, в качестве основных преимуществ использования ExDS9100 можно выделить также:

- уменьшение потребляемой мощности и затрат на охлаждение за счет увеличения аппаратной плотности;
- минимизацию капитальных затрат на оборудование за счет снижения стоимости за терабайт и за терабайт, размещаемый в единице объема;
- упрощение развертывания систем петабайтной емкости с помощью единого графического интерфейса;
- снижение сложности ПО.

Примерами отраслевых применений ExDS9100 могут являться:

- компании, которым требуется хранить и анализировать большое число медицинских снимков;
- центры, связанные с исследованием генома человека;
- охранные агентства, которым требуется хранить большое число роликов с наблюдениями;
- центры обработки геофизической информации (нефть и газ);
- web 2.0 компании, поддерживающие социальные онлайн-порталы и программы по совместному использованию контента;
- компании, связанные с цифровым медиабизнесом: фотосервис, сервис для потокового видео/музыки/анимации.

HP 4400 Scalable NAS File Services

Решение ExDS9100 как и 4400 построено на технологии HP Scalable NAS и разрабатывалось для консолидации отдельных файловых серверов с возможностью также и консолидации блокового хранения в "средних" компаниях, но с учетом всех основных современных тенденций на рынке.

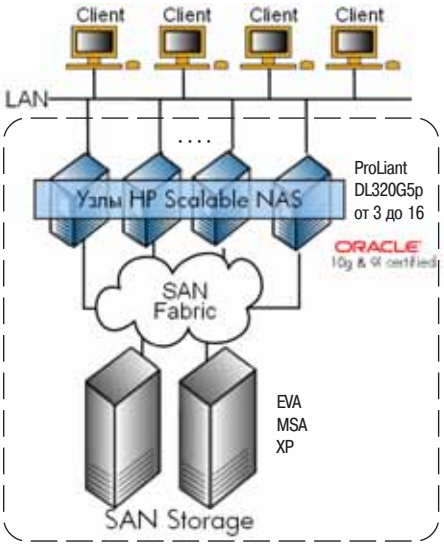

Рис. 4. Структурная схема системы HP Scalable NAS File Services.

Архитектурно система 4400 идентична ExDS9100, но вместо блэйд-серверов используются серверы серии DL (которые называются File Server nodes – FSN) – DL380G5 Clustered Gateway servers, а вместо блоков хранения – FC системы хранения, которые коммутируются с FSN посредством SAN-коммутатора (рис. 4). Поскольку FC системы хранения (в базовой комплектации это EVA) имеют достаточно развитый функционал в составе массива с точки зрения доступности данных (включая ПО репликации и управления и отсутствующее в ExDS9100), а EVA вследствие своих архитектурных особенностей – и максимальное выравнивание нагрузки между LUN, то в целом это дает повышенные показатели по поддержанию уровня доступности и целостности данных, а также предоставляет возможность использования встроенного функционала EVA для управления и репликации данных. Но вместе с этим – и потерю преимуществ по плотности компоновки и большую удельную стоимость единицы хранимых данных.

В базовой конфигурации производительность системы HP 4400 Scalable NAS File Services растет практически линейно как на CIFS, так и на NFS при увеличении числа узлов в системе (рис. 5).

Благодаря виртуализации подключений (так же, как и в Ex1DS9100) все FSN получают доступ ко всему хранилищу данных (HP называет это консолидированной "shared data" архитектурой) и

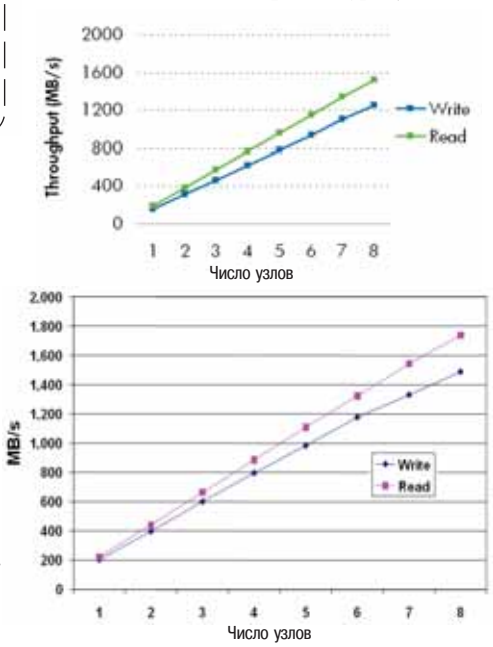

Рис. 5. Базовая конфигурация системы HP 4400 Scalable NAS File Services позволяет практически линейно увеличивать ее производительность как на CIFS (вверху), так и на NFS (внизу) при добавлении узлов.

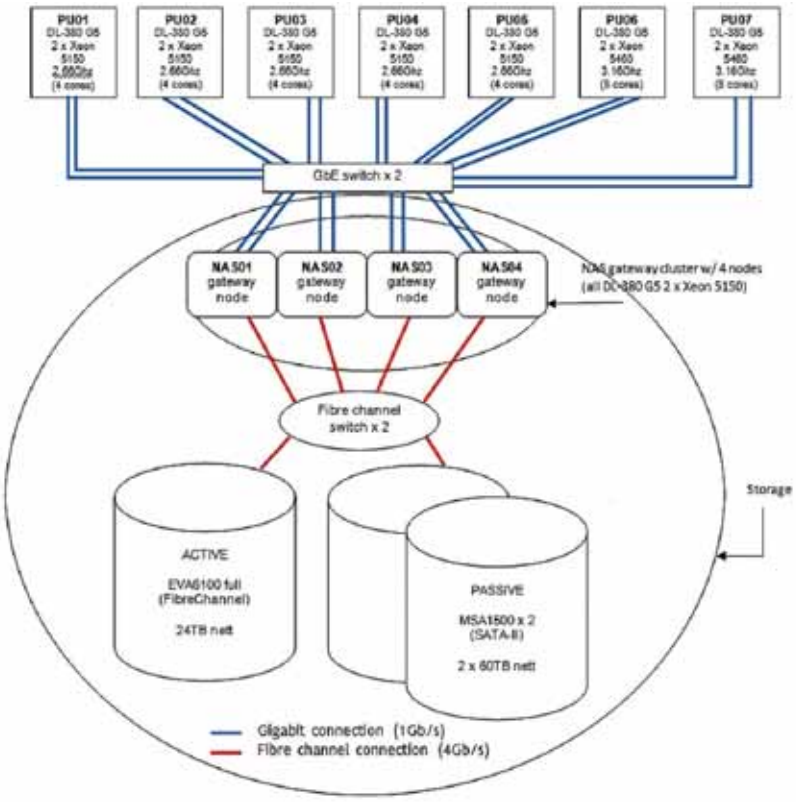

Рис. 6. Конфигурация HP Scalable NAS, развернутая в компании Aerodata.

вместе с хранилищем образуют гибкий динамичный комплекс ресурсов, который можно эффективно использовать тогда, когда это необходимо. Приложения, расположенные на разных серверах, могут получить доступ к одному и тому же пулу данных, что делает возможным перемещение приложения с одного сервера на другой, например, для повышения производительности.

Файловые узлы реализованы на базе ProLiant DL320G5p сервера, которые могут масштабироваться от 3 до 16.

Полностью симметричная архитектура файловой системы позволяет всем узлам "видеть" все файловые системы, что дает возможность избегать узких мест в системе. На файловых узлах могут выполняться различные файловые сервисы и приложения. Всего поддерживается 512 Linux и 256 Windows файловых систем с максимальным размером одной – 128 Тбайт (Linux) и 32 Тбайт (Windows). Максимальная емкость системы – 96 Тбайт.

Поддерживаются протоколы NFS, CIFS, HTTP, FTP, HTTPS, iSCSI. Для хостов с блоковым доступом к данным поддерживаются ОС: HP-UX, HP OpenVMS, Windows 2008/2003 Professional/2003 Standard/Enterprise, Sun Solaris, Linux, IBM AIX, Novell NetWare, VMware и Apple Mac OS X.

По данным HP, архитектура "shared data" позволяет снижать стоимость гигабайта информации более чем на 50% в сравнении с традиционными корпоративными файловыми дисковыми массивами.

Система 4400 позиционируется для "средних" компаний, которым требуется высокая доступность и управляемость, и поставляется полностью предустановленной и готовой для подключения в IP-сеть.

Один из первых проектов на базе HP Scalable NAS был реализован в компании Aerodata, упоминавшейся выше.

Компания Aerodata, занимающаяся аэрофотосъемкой, все свои снимки, как первичные, так и обработанные хранила на USB устройствах – всего около 500. В 2007 г. проблемы, вызванные низкой эффективностью управления большим массивом данных (как в части временных издержек, так и стоимостных), заставили компанию искать более современные технологии. В результате, после анализа четырех вариантов реализации, было выбрано решение на базе HP Scalable NAS, которое было развернуто в сентябре 2007 г. и состояло из HP StorageWorks Enterprise File Services Clustered Gateway, HP StorageWorks 6100 Enterprise Virtual Array (EVA6100) с Fibre Channel дисками общей емкостью 16 Тбайт и HP StorageWorks 1500 Modular Smart Array с SATA-дисками емкостью 60 Тбайт. В настоящее время после добавления емкости ее общий объем составил 150 Тбайт (рис. 6).

Сделанные снимки первоначально сохранялись в пассивном хранилище – MSA1500. Для обработки они перемещались на высокопроизводительную СХД – EVA6100, которая была связана с восьмью HP ProLiant DL380 серверами (четырьмя или восьми ядерными). После обработки данные записывались на USB-устройства для доставки клиентам и перемещались обратно на MSA.

В результате внедрения данного решения удалось в среднем уменьшить время обработки снимков по отдельным проектам до 48 часов, вместо недель и месяцев, как это требовалось раньше. В целом, окупаемость проекта составила всего 7 месяцев, а показатель ROI (Return Of Investments) из расчета на трехлетний период – 222%.

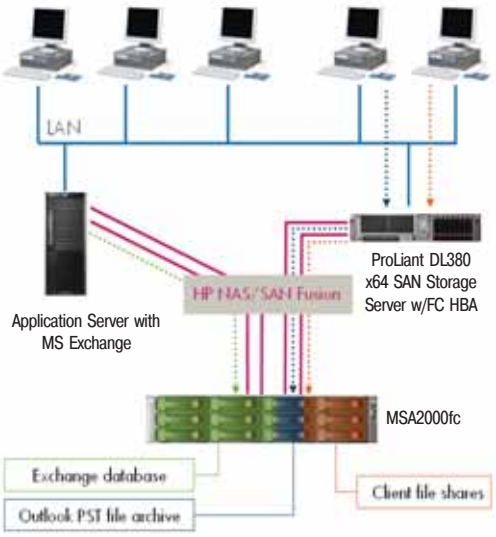

Рис. 8. Пример организации файлового доступа к массиву MSA2000 с помощью ProLiant DL380 x64 SAN Gateway без дорогостоящих FC HBA.

NAS-решения для SMB-рынка и "средних" компаний

Для компаний, работающих в SMB-секторе рынка и "средних" компаний с числом сотрудников/клиентов от 1 до 1000 наиболее приемлемым решением остается уже упомянутый в начале сетевой файловый сервер на базе стандартного под управлением Microsoft Windows Storage Server 2003 R2 (WSS2003 R2) или Windows Unified Data Storage Server 2003 (WUDSS 2003). В настоящее время эти две платформы являются наиболее распространенными для организации унифицированных хранилищ и поставляются только OEM-партнерами Microsoft в составе своих решений.

Из опций, которые были добавлены в WUDSS 2003 в сравнении с WSS2003 R2, можно отметить следующие:

- Single Instance Storage (SIS) – устранение дублирования файлов. Когда SIS находит идентичные файлы, она сохраняет только одну копию, называемую SIS Common Store, все другие файлы замещаются ссылками на нее. Процесс полностью прозрачен для пользователей;
- поддержка полнотекстного индексного поиска для клиентов, работающих под управлением Microsoft Windows 2000/XP. Индекс может включать как ключевые слова, фразы или свойства (например, автор документа) документа;
- поддержка Microsoft iSCSI Software Target, что дает возможность организации блокового доступа через iSCSI-протокол, а также все возможности по управлению самим устройством: создание, управление томами, PKB-сервисами на основе снапшотов;
- поддержка лицензии Microsoft Cluster Server, что дает возможность организации томов более чем 2 Тбайт в кластере.

Возможны 2 варианта реализации NAS-хранилища: в качестве непосредственно файлового хранилища или шлюза (Gateway) к SAN для предоставления файлового доступа (рис. 7). В зависимости от потребностей в последнем случае NAS-решение может комплектоваться различными типами дисков и иметь разную степень масштабируемости.

Пример использования ProLiant DL380 x64 SAN Gateway Storage Server для организации файлового доступа к массиву MSA2000 без дорогостоящих FC HBA дан на рис. 8.

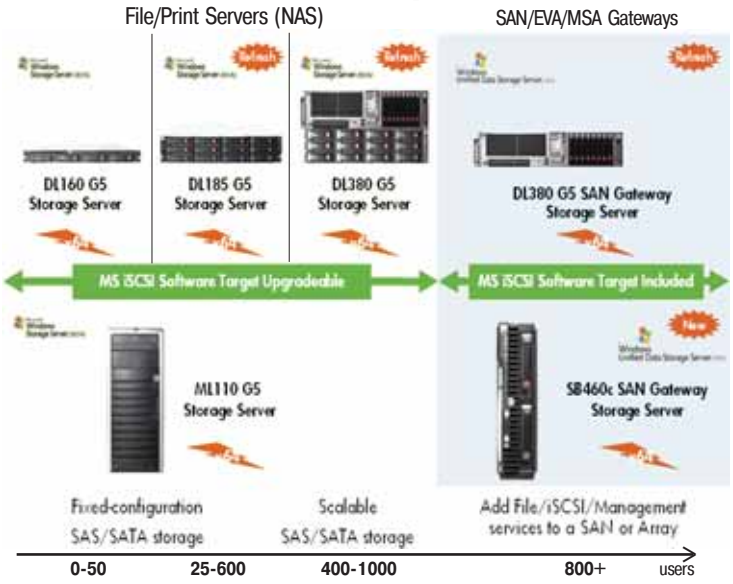

Рис. 7. Позиционирование NAS-решений на основе файлового сервера.

Заключение

Появление в портфеле предложений HP NAS-решений с оптимизированной и с повышенной гибкостью архитектурой на базе технологии HP Scalable NAS позволяет многим компаниям в условиях возрастающих объемов информации и ограничений на IT-бюджет успешно удовлетворять потребности бизнеса в IT-ресурсах, не только не снижая, но и значительно повышая его потенциал.